

UNITED STATES DISTRICT COURT  
DISTRICT OF MASSACHUSETTS

SINGULAR COMPUTING LLC,

Plaintiff,

v.

GOOGLE LLC,

Defendant.

Civil Action No. 1:19-cv-12551-FDS

Hon. F. Dennis Saylor IV

ORAL ARGUMENT REQUESTED

**REDACTED**

**PLAINTIFF'S MOTION IN LIMINE NO. 7**  
**TO PRECLUDE DEFENDANT FROM OFFERING EVIDENCE OR ARGUMENT**  
**CONCERNING ALTERNATIVE CHIPS USING THE NUMBER FORMAT FP16**

Matthew D. Vella (BBO #660171)  
Adam R. Doherty (BBO #669499)  
Kevin Gannon (BBO #640931)  
Brian Seeve (BBO #670455)  
Daniel McGonagle (BBO #690084)  
PRINCE LOBEL TYE LLP  
One International Place, Suite 3700  
Boston, MA 02110  
Tel: (617) 456-8000  
Email: mvella@princelobel.com  
Email: adoherthy@princelobel.com  
Email: kgannon@princelobel.com  
Email: bseeve@princelobel.com  
Email: dmgonagle@princelobel.com

Kerry L. Timbers (BBO #552293)  
SUNSTEIN LLP  
100 High Street  
Boston, MA 02110  
Tel: (617) 443-9292  
Email: ktimbers@sunsteinlaw.com

ATTORNEYS FOR THE PLAINTIFF

At trial, Google plans to present evidence and argument that it could have avoided infringement by using an alternative number format, floatingpoint16 (“fp16”), on either modified tensor processing units (“TPUs”) like the Accused Products or on graphics processing units manufactured by Nvidia (“GPUs”) and that the reasonable royalty due Singular should be limited to the cost of implementing that non-infringing alternative. But the record evidence, including the reports and testimony of Google’s experts, establishes that Google cannot satisfy its burden to show that fp16-based TPUs or fp16-based GPUs were “readily available” at the time of the hypothetical negotiation as required by Federal Circuit precedent.

Based on her discussions with Google engineers, Google’s own damages expert, Laura Stamm, concludes that before fp16 could be used on either modified TPUs or GPUs to achieve the same or similar performance on machine learning tasks as Google achieves with the Accused Products, [REDACTED]

[REDACTED] which they had not performed or begun to perform as of the March 2017 hypothetical negotiation. Under applicable Federal Circuit precedent, a delay of [REDACTED] is far too long for a substitute to be considered a “readily available” non-infringing alternative. *See e.g., Micro Chemical, Inc. v. Lextron, Inc.*, 318 F.3d 1119, 1123 (Fed. Cir. 2003) (reversing summary judgment finding that alternative that required months of design and testing was readily available).

To avoid this binding, issue-determinative case law, Google and Ms. Stamm plan to attempt an end-run around the law of “availability” in two ways. First, as Ms. Stamm explains in her expert report and deposition testimony, they will argue that Ms. Stamm does not rely on modified TPUs or GPUs using fp16 as a non-infringing alternative and, as a result, the availability requirement does not apply. But there is no escaping the fact that she relies on TPUs

or GPUs using fp16 as a non-infringing alternative. Ms. Stamm analyzes the increased cost of implementing those chips using fp16 instead of the number format used by the Accused Products, bf16, and opines that the reasonable royalty would be no more than those increased costs, ██████████ While Ms. Stamm's opinion is erroneous for other reasons, her general approach is consistent with the well accepted cost-savings methodology. *See e.g., Prism Technologies LLC v. Sprint Spectrum L.P.*, 849 F.3d 1360, 1376 ("A price for a hypothetical license may appropriately be based on the 'costs and availability of non-infringing alternatives' and the potential infringer's 'cost savings.'"). The availability requirement applies and Google cannot satisfy it.

It also appears Ms. Stamm and Google will attempt to avoid the fact that TPUs and GPUs using fp16 were not available by arguing that availability should be judged as of the date Google began development of the Accused Products (i.e. ██████████ prior to the hypothetical negotiation), not as of the date of the hypothetical negotiation. The law is clear, however, that a reasonable royalty determination must be made based on the facts as they existed at the time of the hypothetical negotiation and not based on an after-the-fact assessment of what the accused infringer arguably could have done, but did not do. As Ms. Stamm admitted at her deposition, when analyzed based on the facts as they existed at the time of the March 2017, modified TPUs or GPUs using fp16 would not have been available until ██████████ at the earliest. That admission is determinative.

For these reasons and the reasons set forth below, Singular respectfully requests that the Court preclude Google from presenting (1) evidence or argument that a TPU or GPU using fp16 was an available non-infringing alternative at the time of the hypothetical negotiation and (2)

evidence or argument concerning the costs associated with designing, developing, deploying or using TPUs or GPUs running fp16 instead of bf16.

## I. FACTUAL AND PROCEDURAL BACKGROUND

### A. Google's Reliance on Non-Infringing Alternatives

Through Ms. Stamm and its technical expert, Dr. Martin Walker, Google presents two non-infringing alternatives on which it relies in an attempt to limit Singular's damages, both of which are alternatives to using brainfloat16 ("bf16"), the number format used by the Accused Products, TPU v.2 (a.k.a. [REDACTED] and TPU v.3 (a.k.a. [REDACTED]

Google primarily relies on a number format called brainfloat20 ("bf20") to argue that Singular's reliance on Nvidia GPUs as the next-best available non-infringing alternative was erroneous, because, at the time of the hypothetical negotiation in March 2017, Google could have avoided infringement by deploying modified TPUs using bf20 instead of bf16. Ex. A,

Stamm Rpt. at ¶ 15 [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

(bold emphasis added); Ex. B, Stamm Tr. at 62 ("I believe the bfloat20 is the next best alternative."). Singular's motion to exclude all evidence concerning bf20 on the grounds that Google failed to timely disclose it and that it was not available at the time of the hypothetical negotiation is currently pending before the Court. *See* ECF No. 466.

Alternatively, Google relies on either modified TPUs or GPUs using fp16 instead of bf16. Like the implementation of a chip using the alleged alternative bf20 number format, Ms. Stamm concedes that it would cost Google more to design, develop and operate a chip using fp16 than it cost Google to design, develop and use bf16 on the Accused Products. [REDACTED]

[REDACTED]  
[REDACTED]  
[REDACTED]  
[REDACTED]  
[REDACTED]  
[REDACTED]  
[REDACTED]

Based on her discussions with Google engineers, Ms. Stamm explains that the work necessary to perform the loss scaling would have taken [REDACTED] Google engineers at least [REDACTED] to complete. *Id.* at ¶ 171 [REDACTED]

[REDACTED]  
[REDACTED]  
[REDACTED]  
[REDACTED] *Id.* at ¶ 171 (emphasis added). According to Ms. Stamm, additional work would be required after the initial [REDACTED] period before the chips could be deployed. *Id.* at ¶ 172 [REDACTED]

[REDACTED]  
[REDACTED]  
Ms. Stamm's opinions with respect to fp16 and the work necessary to make it suitable for the machine learning tasks for which Google uses the Accused Products are not based on the time of the hypothetical negotiation. Instead, Ms. Stamm states that it is the work Google would have had to start in [REDACTED] in order for fp16 to be available at the time of the hypothetical negotiation in March 2017. *Id.* at ¶ 171 [REDACTED]

[REDACTED] (bold emphasis added). At her deposition, Ms. Stamm testified that the

time for Google to evaluate alternatives was not the time of the hypothetical negotiation in March 2017, it was [REDACTED] when they began development of the Accused Products. Ex. B., Stamm Tr. at 155 [REDACTED]  
[REDACTED]  
[REDACTED]

Ultimately, Ms. Stamm opines that the cost to perform the loss scaling, provide the post-loss scaling tools and automation, and to provide support for the chips while they were deployed would have cost Google [REDACTED] million. Ex. A, Stamm Rpt. at ¶ 174, Table 4. After combining those costs with the additional electricity expense that would have been incurred using fp16, Ms. Stamm opines that the reasonable royalty would not have exceeded \$38 million. *Id.*

B. Google’s Concession That a Non-Infringing Alternative Must Be “Available” at the Time of the Hypothetical Negotiation.

In its opposition to Singular’s motion to exclude evidence concerning bf20, Google concedes that, in order for bf20 to serve as a non-infringing alternative, the law requires that it be “available” at the time of the hypothetical negotiation. *See* ECF No. 509 at § III.B. (“The bfloat20 alternative was available at the time of the hypothetical negotiation.”); *id.* at 14-15 (relying on *Grain Processing Corp. v. Am. Maize-Prod. Co.*, 185 F.3d 1341, 1349 (Fed. Cir. 1999) to argue that, although bf20 was not on the market at the time of the hypothetical negotiation, it is still available under the law because the evidence is sufficient to show Google “could readily obtain all of the materials needed” to produce bf20, it “had all of the necessary equipment, know-how, and experience” to make the substitute, and making bf20 was “not prohibitively expensive.”).

C. Ms. Stamm's Contention That The "Availability" Requirement Does Not Apply To Google's Reliance on TPUs or GPUs Using fp16 As An Alternative to the Accused Products Using bf16.

At his deposition, Google's technical expert,

Notwithstanding Dr. Walker's testimony

her comparison of the Accused Products to

[REDACTED]

[REDACTED]

## II. ARGUMENT

### A. The Law of “Availability” Applies to Google’s Reliance on GPUs Using fp16 As An Alternative to the Accused Products Using bf16.

Reliance on non-infringing alternatives to isolate the benefit of an invention is a well-accepted methodology for determining a reasonable royalty arising out of a hypothetical negotiation. *Prism Technologies LLC v. Sprint Spectrum L.P.*, 849 F.3d 1360, 1376 (the value of patented technology can properly be valued “show[ing] that the defendants’ infringement allowed it to avoid taking a different, more costly course of action. A price of a hypothetical license may appropriately be based on consideration of the ‘costs and availability of non-infringing alternatives’ and the potential infringer’s ‘cost savings.’”). To employ this methodology as both parties do in this case, the proponent of the non-infringing alternative must show that the alternative was on the market or “available” at the time of infringement. *Grain Processing Corp. v. American Maize-Products Co.*, 185 F.3d 1341, 1350 (Fed. Cir. 1999) (internal citations omitted). Where the accused infringer relies on an alleged alternative that was not on the market at the time of the hypothetical negotiation, it bears the burden to establish that, at the time of infringement (i.e. at the time of the hypothetical negotiation), it had the “necessary equipment, know-how, and experience” to “readily” implement the non-infringing alternative. *Grain Processing*, 185 F.3d at 1354-55; see also *Micro Chemical, Inc. v. Lextron, Inc.*, 318 F.3d 1119, 1122-23 (Fed. Cir. 2003) (clarifying the *Grain Processing* standard by explaining that the accused infringer must establish that its “equipment, know-how, and experience” would have

allowed it to “readily” use the non-infringing alternative);<sup>1</sup> *see also Stoller Enterprises, Inc. v. Fine Agrochemicals Ltd.*, No. 4:20-cv-00750, 2023 WL 8283633, at \*16 (S.D. Tex. Nov. 30, 2023) (holding that “it is Defendants’ burden to establish that ‘any alternatives [were] acceptable and available’ under reasonable royalties analysis in order to get the benefit of a lower calculation.”) (internal citations omitted).

Google previously acknowledged that it may only rely on TPUs using bf20 as a non-infringing alternative if it can establish that its “equipment, know-how, and experience” would have allowed it to deploy those chips at the time of the hypothetical negotiation. *See* ECF No. 509 at § III.B. It appears, however, that with regard to its second alternative, modified TPUs or GPUs using fp16, it and its damages expert, Ms. Stamm, [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED] But that is exactly what the non-infringing alternative methodology is designed to do; it is intended to determine the value of the invention

---

<sup>1</sup> As explained in Singular’s pending motion regarding the alleged bf20 alternative, ECF No. 466-1 at 6, *Grain Processing* is a lost profits case which established that the plaintiff bears the burden to establish the absence of non-infringing alternatives on the market. 185 F.3d at 1351-52. If there is no alternative on the market, the court is free to infer that an alleged alternative not on the market is not available. The defendant can rebut that inference by showing that it had the “necessary equipment, know-how, and experience” to readily implement the non-infringing alternative. *Id.* The Federal Circuit has not addressed whether *Grain Processing* applies in a case involving reasonable royalty damages. *See MicroVention, Inc. v. Balt USA, LLC*, C.A. No.: 19-cv-01335, 2022 WL 18284997 at \*2 (C.D. Cal. Dec. 12, 2022). However, various courts have extended *Grain Processing* to a reasonable royalty analysis. *See LaserDynamics, Inc. v. Quanta Comp., Inc.*, No. 2:06-CV-348, 2011 WL 197869, at \*3 (E.D. Tex. Jan. 20, 2011); *Pavo Sols. LLC v. Kingston Tech. Co.*, No. 8:14-cv-01352-JLS-KES, 2019 WL 8138163, at \*20-21 (C.D. Cal. Nov. 20, 2019); *See SPEX Techs. v. Apricorn, Inc.*, No. CV 16-07349 JVS (AGR), 2020 WL 1289546, at \*2 (C.D. Cal. Jan. 21, 2020); *Smart Skins LLC v. Microsoft Corp.*, No. C15-544-MJP, 2016 WL 4148091, at \*2 (W.D. Wash. July 1, 2016).

based on the cost of implementing a non-infringing alternative. Ms. Stamm's assertions amount to semantics.

There is no difference between Ms. Stamm's comparison of the costs associated with designing, developing and using bf16 on the Accused Products and alternatively using bf20, which Google acknowledges must be "available" to be considered a non-infringing alternative at trial. Based on her discussions with Dr. Walker and Google engineers, Ms. Stamm opined that it would have cost Google an additional [REDACTED] million to build TPUs using bf20 than it cost Google to build the Accused Products using bf16. Ex. A, Stamm Rpt. at ¶ 15. She applies the same methodology to determine the alleged cost savings using fp16 in modified TPUs or GPUs. *Id.* at ¶ 16. There is no difference and neither Ms. Stamm nor Google can explain why the availability requirement applies to the alleged bf20 alternative, but not to the alleged fp16 alternative. It applies and, as explained below, Ms. Stamm's own report and testimony establish that neither TPUs nor GPUs using fp16 were available at the time of the hypothetical negotiation.

B. All Evidence Concerning fp16 Should Be Excluded Because No Reasonable Jury Could Conclude That fp16 Was Available At The Time of the Hypothetical Negotiation In March 2017.

In *Grain Processing*, the Federal Circuit affirmed the trial court's finding of availability where the accused infringer was able to show that it need not design around the patented technology and that, using its equipment, know-how, and experience, it was able to implement the non-infringing alternative in just two weeks, which the court described as "practically instantaneous." *Grain Processing*, 185 F.3d at 1346. The Court reached the opposition conclusion in *Micro Chemical*. In that case, the evidence establishes that the accused infringer expended 984 hours to design the non-infringing alternative, another 330 to test it and that one of its engineers worked on the design full-time for several months. *Micro Chemical*, 318 F.3d at

1123. It then took another two months to convert all of the infringing machines to non-infringing machines. *Id.* Based on those facts, the Federal Circuit concluded that the accused infringer did not have “the necessary equipment, know-how, and experience” to make the non-infringing alternative at the time of infringement. *Id.*

Ms. Stamm’s report and testimony establish that the work necessary to use fp16 on modified TPUs or GPUs was even more extensive than the design and development work in *Micro Chemical*. Based on her discussions with a Google engineer and Google’s Rule 30(b)(6) witness on the topic of non-infringing alternatives, Dr. Nashant Patil,<sup>2</sup> Ms. Stamm concluded that an entire team of [REDACTED] Google engineers would have needed to work for [REDACTED] just to perform the loss scaling work necessary to increase the dynamic range of fp16, Ex. A, Stamm Rpt. at ¶ 171, and after that initial [REDACTED] period, that [REDACTED] member team would have to continue its work for an unspecified amount of time before TPUs or GPUs using fp16 could actually be deployed. *Id.* at ¶ 172. In light of the Federal Circuit’s decision in *Micro Chemical*, there is no reasonable basis to argue that, at the time of the hypothetical negotiation in March 2017, Google had the “necessary equipment, know-how, and experience” to readily use fp16 on TPUs or GPUs. *Micro Chemical*, 318 F.3d at 1123 (rejecting availability argument based on one engineer working full-time for several months); *SynQor, Inc. v. Artesyn Techs., Inc.*, No. 2:07-cv-497, 2011 WL 3625036, at \*10-11 (E.D. Tex. Aug. 17, 2011) (stating that exclusion of non-infringing alternatives on procedural grounds was harmless because the alternative was “more than eight months away from being ready.”).

---

<sup>2</sup> [REDACTED]

[REDACTED]

[REDACTED]

Google and Ms. Stamm implicitly acknowledge this hurdle by ignoring the law and arguing that the availability of fp16 on TPUs or GPUs should be determined based on an after-the-fact assessment of what Google could have done, but did not do, when it first started developing the Accused Products two years prior to the hypothetical negotiation. For example,

Ms. Stamm states [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED] *Id.* at ¶ 171 (emphasis added).

The law requires, however, that the reasonable royalty be based on the facts as they existed at the time of the hypothetical negotiation, which takes place at the time of first infringement. As the Federal Circuit has repeatedly explained, “[a] reasonable royalty determination for purposes of making a damages evaluation must relate to the time infringement occurred, and not be an after-the-fact assessment.” *Riles v. Shell Exploration & Prod. Co.*, 298 F.3d 1302, 1313 (Fed. Cir. 2002); *see also Uniloc USA, Inc. v. Microsoft Corp.*, 632 F.3d 1292, 1318 (Fed. Cir. 2011) (expert opinions must be tied to the facts and “the hypothetical negotiations that would have taken place in light of those facts and circumstances at the relevant time.”); *Hanson v. Alpine Valley Ski Area, Inc.*, 718 F.2d 1075, 1081-82 (Fed. Cir. 1983) (“Having [chosen to use infringing machines rather than a non-infringing alternative, the accused infringer] cannot invalidate an otherwise reasonable royalty on the claim that by hindsight it would have been better off if it had purchased the non-infringing . . . machines.”); *Fresenius Med. Care Holdings, Inc. v. Baxter Int’l, Inc.*, No. C 03-01431 SBA, 2006 WL 1646113, at \*1

(N.D. Cal. June 12, 2006) (“[A] key part of the reasonable royalty determination under *Georgia Pacific* is whether the accused infringer had acceptable non-infringing alternatives available to it at the time of the hypothetical negotiation.”).

It is undisputed that, at the time of the hypothetical negotiation in March 2017, Google had not begun the loss scaling necessary to use fp16 on TPUs or GPUs. The law constrains Google to argue availability based on what it could have done as of the March 2017 hypothetical negotiation and, as even Ms. Stamm admits, if Google started the necessary loss scaling in March 2017, TPUs or GPUs using fp16 would not have been available until [REDACTED] at the earliest. See Ex. B, [REDACTED]  
[REDACTED]  
[REDACTED]

Accordingly, there is no basis on which to allow Google to introduce evidence concerning the use of fp16 on modified TPUs or GPUs and all such evidence should be precluded.

### III. CONCLUSION

For the foregoing reasons, Singular respectfully requests that the Court preclude Google from presenting (1) evidence or argument that a TPU or GPU using fp16 was an available non-infringing alternative at the time of the hypothetical negotiation and (2) evidence or argument concerning the costs associated with designing, developing, deploying or using TPUs or GPUs running fp16 instead of bf16.

Dated: December 5, 2023

Respectfully submitted,

*/s/ Kevin Gannon*

---

Matthew D. Vella (BBO #660171)  
Adam R. Doherty (BBO #669499)  
Kevin Gannon (BBO #640931)  
Brian Seeve (BBO #670455)  
Daniel McGonagle (BBO #690084)  
**PRINCE LOBEL TYE LLP**  
One International Place, Suite 3700  
Boston, MA 02110  
Tel: (617) 456-8000  
Email: mvella@princelobel.com  
Email: adoherthy@princelobel.com  
Email: kgannon@princelobel.com  
Email: bseeve@princelobel.com  
Email: dmgonagle@princelobel.com

Kerry L. Timbers (BBO #552293)  
**SUNSTEIN LLP**  
100 High Street  
Boston, MA 02110  
Tel: (617) 443-9292  
Email: ktimbers@sunsteinlaw.com

ATTORNEYS FOR THE PLAINTIFF

**LOCAL RULE 7.1 CERTIFICATION**

I, Kevin Gannon, counsel for Singular Computing LLC, hereby certify that I conferred with counsel for Google LLC to resolve the issues presented in this motion but, after a good faith attempt to reach agreement, the parties were unable to do so.

*/s/ Kevin Gannon*

**CERTIFICATE OF SERVICE**

I certify that, on December 5, 2023, all counsel of record who have consented to electronic service are being served with a copy of this document via the Court's CM/ECF system.

*/s/ Kevin Gannon*